Lecture 3: Continuous times Markov chains. Poisson Process. Birth and Death process.

Antonina Mitrofanova, NYU, department of Computer Science

December 18, 2007

1 Continuous Time Markov Chains

In this lecture we will discuss Markov Chains in continuous time. Continuous time Markov Chains are used to represent population growth, epidemics, queueing models, reliability of mechanical systems, etc.

In Continuous time Markov Process, the time is perturbed by exponentially distributed *holding times* in each state while the succession of states visited still follows a discrete time Markov chain. Given that the process is in state i, the holding time in that state will be exponentially distributed with some parameter λ_i, where i can represent the current population size, the number of alleles A_1 in the population, etc. These holding times basically control how rapidly the movements (changes of states) of the chain take place. Additionally, given the knowledge of visited states, the holding times are independent random variables.

For a Continuous Markov Chain, the transition probability function for $t > 0$ can be described as

$$P_{ij}(t) = P(X(t+u) = j|X(u) = i)$$

and is independent of $u \geq 0$. In fact, $P(X(t+u) = j|X(u) = i)$ is a function of t and describes a *time-homogeneous* transition law for this process.

To construct a Markov process in discrete time, it was enough to specify a one step transition matrix together with the initial distribution function. However, in continuous-parameter case the situation is more complex. The specification of a single transition matrix $[p_{ij}(t_0)]$ together with the initial distribution is not adequate. This is due to the fact that events that depend on the process at time points that are not multiples of t_0 might be excluded. However, if one specifies all transition matrices $p(t)$ in $0 < t \leq t_0$ for some $t_0 > 0$, all other transition probabilities may be constructed from these. These transition probability matrices should be chosen to satisfy the Chapman-Kolmogorov equation, which states that:

$$P_{ij}(t+s) = \sum_k P_{ik}(t)P_{kj}(s)$$

Or we can state it in a matrix notation by the following so-called *semigroup property*:

$$\mathbf{P}(t+s) = \mathbf{P}(t)\mathbf{P}(s)$$

The (i,j) element of the matrix $\mathbf{P}(t+s)$ is constructed by i row of $\mathbf{P}(t)$ multiplied by the j column of $\mathbf{P}(s)$. For some time points $0 < t_1 < t_2 < t_3$ and arbitrary states a, b, c, d one has that:

$$P(X_0 = a, X_{t_1} = b, X_{t_2} = c, X_{t_3} = d) = p_a p_{a,b}(t_1) p_{b,c}(t_2 - t_1) p_{c,d}(t_3 - t_2)$$

as well as

$$P(X_0 = a, X_{t_1} = b, X_{t_3} = d) = p_a p_{a,b}(t_1) p_{c,d}(t_3 - t_1)$$

The consistency of the Chapman-Kolmogorov Equation would require the following:

$$p_{c,d}(t_3 - t_1) = \sum_c p_{b,c}(t_2 - t_1) p_{c,d}(t_3 - t_2)$$

The above postulates give motivation for *Kolmogorov forward and backward equations*, which will be discussed in later sections in detail. Let us start with introduction of a Continious time Markov Chain called *Birth-and-Death process*.

2 Birth-and-Death process: An Introduction

The birth-death process is a special case of continuous time Markov process, where the states (for example) represent a current size of a population and the transitions are limited to birth and death. When a birth occurs, the process goes from state i to state $i+1$. Similarly, when death occurs, the process goes from state i to state $i-1$. It is assumed that the birth and death events are independent of each other.

The birth-and-death process is characterized by the birth rate $\{\lambda_i\}_{i=0,\ldots,\infty}$ and death rate $\{\mu_i\}_{i=0,\ldots,\infty}$, which vary according to state i of the system. We can define a *Pure Birth Process* as a birth-death process with $\mu_i = 0$ for all i. Similarly, a *Pure Death Process* corresponds to a birth-death process with $\lambda_i = 0$ for all i.

The general description of the Birth-and-Death process can be as follows: after the process enters state i, it holds (sojourns) in the given state for some random length of time, exponentially distributed with parameter $(\lambda_i + \mu_i)$. When leaving i, the process enters either $i+1$ with probability

$$\frac{\lambda_i}{\lambda_i + \mu_i}$$

or $i-1$ with probability

$$\frac{\mu_i}{\lambda_i + \mu_i}$$

If the next state chosen is $i+1$, then the process sojourns in this state according to the exponential distribution with parameter $\lambda_{i+1} + \mu_{i+1}$ and then chooses the next state etc. The number of visits back to the same state is ignored since in a continuous time process transitions from state i back to i would not be identifiable.

Imagine having two exponentially distributed random variables $B(i)$ and $D(i)$ with parameters λ_i and μ_i respectively. These random variables describe the holding time in the state i. We can think of $B(i)$ as the time until a birth and $D(i)$ is the time until a death (when a population size is i). The population increases by one if the birth occurs prior to death and decreases by one otherwise. If $B(i), D(i)$ are independent exponentially distributed random variables, then their minimum is exponentially distributed with parameter $(\lambda_i + \mu_i)$.

A transition from i to $i+1$ is made if $B(i) < D(i)$, which occurs with probability

$$P[B(i) < D(i)] = \frac{\lambda_i}{\lambda_i + \mu_i}$$

This motion is analogous to a random walk with the difference that here the transitions occur at random times (as opposed to fixed time periods in random walks).

It is of necessity to discuss the Poisson process, which is a cornerstone of stochastic modelling, prior to modelling birth-and-death process as a continuous Markov Chain in detail.

2.1 The law of Rare Events

The common occurrence of Poisson distribution in nature is explained by *the law of rare events*. Consider a large number N of independent Bernoulli trials where the probability p of success on each trial is small and constant from trial to trial. Let $X_{N,p}$ be the total number of successes in N trials, where $X_{N,p}$ follows the binomial distribution, for $k = 0, 1, \ldots, N$.

$$P(X_{N,p} = k) = \binom{N}{k} p^k (1-p)^{N-k}$$

If we assume that $N \longrightarrow \infty$ and $p \longrightarrow 0$, so that $Np = \mu$, then the distribution for $X_{N,p}$ becomes the Poisson distribution:

$$P(X_\mu = k) = \frac{e^{-\mu}\mu^k}{k!} \qquad for\ k = 0, 1, \ldots$$

In Stochastic modelling, this law is used to suggest circumstances under which the poisson distribution might be expected to prevail, at least approximately.

2.2 Poisson Process

A poisson distribution with parameter $\mu > 0$ is given by

$$p_k = \frac{e^{-\mu}\mu^k}{k!}$$

and describes the probability of having k events over a time period embedded in μ. The random variable X having a Poisson distribution has the mean $E[X] = \mu$ and the variance $Var[X] = \mu$.

The Poisson process entails notions of Poisson distribution together with independence. A Poisson process of intensity $\lambda > 0$ (that describes the expected number of events per unit of time) is an integer-valued Stochastic process $\{X(t); t \geq 0\}$ for which:

1. for any arbitrary time points $t_0 < t_1 < t_2 < \cdots < t_n$, and $t_0 = 0$, the number of events happening in disjoint intervals (process increments)

   $$X(t_1) - X(t_0), X(t_2) - X(t_1), X(t_3) - X(t_2), \dots, X(t_n) - X(t_{n-1})$$

   are independent random variables. This means that the number of events in one time interval is independent from the number of events in an interval that is disjoint from the first interval. This is known as *independent increments* property of the Poisson process.

2. for $s \geq 0$ and $t > 0$, the random variable $X(s+t) - X(s)$, which describes the number of events occurring between time s and $s + t$ (independent increment), follows the Poisson distribution

   $$P(X(s+t) - X(s) = k) = \frac{(\lambda t)^k e^{-\lambda t}}{k!}$$

3. We assume that at time zero the number of events that have happened already is zero.

In this case, the parameter of the Poisson distribution is λt, $E[X(t)] = \lambda t$, and $Var[X(t)] = \lambda t$. Let us fix a short interval of time h. In Stochastic modelling, it is of a special interest to derive the probability of exactly one event over the time period h:

$$\begin{aligned}
P(X(t+h) - X(t) = 1) &= \frac{(\lambda h)e^{-\lambda h}}{1!} && (1)\\
&= (\lambda h)\sum_{n=0}^{\infty}\frac{(-\lambda h)^n}{n!} && (2)\\
&= (\lambda h)(\frac{(-\lambda h)^0}{0!} + \frac{(-\lambda h)^1}{1!} + \frac{(-\lambda h)^2}{2!} + \dots) && (3)\\
&= (\lambda h)(1 - \lambda h + \frac{1}{2}\lambda^2 h^2 - \dots) && (4)\\
&= \lambda h + o(h) && (5)\\
& && (6)
\end{aligned}$$

where $o(h)$ denotes a general and unspecified remainder term of smaller order than h. We can view the rate λ in Poisson process $X(t)$ as the proportionality constant in the probability of an event occurring during an arbitrary small interval h.

In a Poisson Process, the waiting time between consecutive events is called a *Sojourn time*, $S_i = W_{i+1} - W_i$, where W_i is the time of occurrence of the i'th event. Basically, S_i measures the duration that the Poisson process sojourns in state i.

The Sojourn times $S_0, S_1, \dots, S_{n-1}$ are independent random variables, each having the exponential probability density function

$$f_{S_k}(s) = \lambda e^{-\lambda s}$$

2.3 Definition of Birth-and-Death process

We let $\{X(t)\}_{t\geq 0}$ be a Markov chain and define a very short interval of time $h \searrow 0$, during which there exist observable changes in a chain. We would like to calculate a probability of seeing some particular changes occurring at time $t+h$, given that we started at time t. Over such a short interval of time $h \searrow 0$, it is nearly impossible to observe more than one event; in fact, the probability to see more than one event is $o(h)$.

If we are to describe a pure birth process with the birth rate λ_i, we would name it a Poisson process with parameter $\lambda_i h$ so that λ_i is the expected number of birth events that occur per unit time. In this case, the probability of a birth over a short interval h is $\lambda_i h + o(h)$.

Similarly, if in state $X(t) = i$ a death rate is μ_i, then the probability that an individual dies in a very small time interval of length h is $\mu_i h + o(h)$.

In the case of birth-and-death process, we have both birth and death events possible, with rates λ_i and μ_i accordingly. Since birth and death processes are independent and have poisson distribution with parameters $\lambda_i h$ and $\mu_i h$, their sum is a Poisson distribution with parameter $h(\lambda_i + \mu_i)$.

Let us analyze changes which might occur in birth-and-death process over the time interval h. Given that there are currently i people in the population, as $h \searrow 0$, the probability that there will be an change of size 1 is basically represented by the probability of one birth and no death (which is a main probabilitic component) or other combinations (two birth and one death, three birth and two death, etc) chances of which however are very small, $o(h)$:

$$
\begin{aligned}
P_{i,i+1}(h) &= P(X(t+h) - X(t) = 1 | X(t) = i) && (7)\\
&= \frac{(\lambda_i h)^1 e^{-\lambda_i h}}{1!} \frac{(\mu_i h)^0 e^{-\mu_i h}}{0!} + o(h) && (8)\\
&= (\lambda_i h) e^{-\lambda_i h} e^{-\mu_i h} + o(h) && (9)\\
&= (\lambda_i h) e^{-h(\lambda_i + \mu_i)} && (10)\\
&= (\lambda_i h) \sum_{n=0}^{\infty} \frac{(-h(\lambda_i + \mu_i))^n}{n!} && (11)\\
&= (\lambda_i h)(1 - h(\lambda_i + \mu_i) + \frac{1}{2} h^2 (\lambda_i + \mu_i)^2 - \dots) + o(h) && (12)\\
&= \lambda_i h + o(h) && (13)
\end{aligned}
$$

In this case, $o(h)$ term represents the fact that there are two birth and one death, 3 birth and 2 death, etc . As h gets really, the probability that $o(h)$ possibilities occur vanishes.

Similarly, the probability that there will be a decreasing change of size 1 is

$$P_{i,i-1}(h) = P(X(t+h) - X(t) = -1 | X(t) = i) = \mu_i h + o(h)$$

Basically, the above postulates assume that the probabilities of population increasing or decreasing by 1 are proportional to the length of the interval. In general, the process is called a birth-and-death process if:

(1) $P(X(t+h) - X(t) = 1 | X(t) = i) = \lambda_i h + o(h)$

(2) $P(X(t+h) - X(t) = -1 | X(t) = i) = \mu_i h + o(h)$

(3) $P(|X(t+h) - X(t)| > 1 | X(t) = i) = o(h)$

(4) $\mu_0 = 0, \lambda_0 > 0; \quad \mu_i, \lambda_i > 0, i = 1, 2, 3, \dots$

These postulates support the notion that the events are rare and almost exclude the possibility of simultaneous occurrence of two or more events. Basically, only one event can occur in a very small interval of time h. And even though the probability for more than one event is non-zero, it is negligible.

The above implies that

(5) $P(X(t+h) - X(t) = 0 | X(t) = i) = 1 - (\mu_i + \lambda_i)h + o(h)$

We will postulate $P_{ij}(h)$ for h small and then derive a system of differential equations satisfied by $P_{ij}(t)$ for all $t > 0$.

3 Sojourn times

Let S_i be a random variable describing a sojourn time of $X(t)$ in state i. That is, given that a process is in state i, what is the distribution of the time S_i the process first leaves state i.

Let us define $G_i(t)$:

$$G_i(t) = P(S_i \geq t)$$

Then by the Markov property it follows that as $h \searrow 0$,

$$\begin{aligned} G_i(t+h) &= G_i(t)G_i(h) = G_i(t)[P_{ii}(h) + o(h)] && (14) \\ &= G_i(t)[1 - h(\lambda_i + \mu_i)] + o(h) && (15) \\ &= G_i(t) - G_i(t)h(\lambda_i + \mu_i) + o(h) && (16) \end{aligned}$$

Subtracting $G_i(t)$ from both sides and dividing by h gives

$$\begin{aligned} \frac{G_i(t+h) - G_i(t)}{h} &= \frac{G_i(t) - G_i(t)h(\lambda_i + \mu_i) + o(h) - G_i(t)}{h} && (17) \\ \frac{G_i(t+h) - G_i(t)}{h} &= -G_i(t)(\lambda_i + \mu_i)] + o(1) && (18) \\ G_i^{'}(t) &= -G_i(t)(\lambda_i + \mu_i) && (19) \end{aligned}$$

We can solve the above by applying the fact that if $y^{'}(x) = Ay(x)$, then $y(x) = e^{xA}$ [1]. Using the condition $G_i(0) = 1$, the solution to the equations is

$$G_i(t) = e^{-t(\lambda_i + \mu_i)}$$

Using the facts that $G_i(x) = 1 - P(S_i \leq x)$ and the cumulative distribution function of the exponential distribution is $P(X \leq x) = 1 - e^{-\lambda x}$, we conclude that S_i follows exponential distribution with parameter $(\lambda_i + \mu_i)$ and mean (expectation) $1/\lambda_i + \mu_i$.

4 Infinitesimal Generator of the Birth-and-Death process

The birth-and-death process is defined as

$$p_{ij}(h) = \begin{cases} \lambda_i h + o(h), & \text{if j=i+1} \\ \mu_i h + o(h), & \text{if j=i-1} \\ 1 - h(\lambda_i + \mu_i) + o(h), & \text{if j=i} \\ o(h), & \text{otherwise} \end{cases}$$

We can condense this notation by writing

$$p_{ij}(h) = \delta_{ij} + hq_{ij} + o(h)$$

where

$$\delta_{ij} = \begin{cases} 1, & j = i \\ 0, & j \neq i \end{cases} \qquad q_{ij} = \begin{cases} \lambda_i, & \text{if j=i+1} \\ \mu_i, & \text{if j=i-1} \\ -(\lambda_i + \mu_i), & \text{if j=i} \\ 0, & \text{otherwise} \end{cases}$$

The δ_{ij} is called a Kronecker's delta, $\delta_{ij} = \lim_{t \downarrow 0} p_{i,j}(t)$. It is given by $\delta_{ij} = 1$ if $i = j$ and $\delta_{ij} = 0$ if $i \neq j$. This condition is reasonable in most circumstances: it requires that with probability one the process spends a

[1] Since $\frac{y^{'}(x)}{y(x)} = A$, it implies that $\frac{d}{dt} ln(y(x)) = A$. Integrating both sides gives $ln(y(x)) = Ax + c$. Now, $y(x) = e^{Ax+c} = e^c e^{Ax} = ae^{Ax}$.

possible (but variable) amount of time in the initial state i before moving to a different state j. This relation can also be expressed in matrix notation

$$\lim_{t \downarrow 0} \mathbf{p}(t) = \mathbf{I}$$

where **I** is the identity matrix (with 1's along the diagonals and 0's elsewhere). We shall also write

$$\mathbf{p}(0) = \mathbf{I}$$

The q_{ij} are called *transition rates*, and $[q_{ij}]$ define the matrix **Q**, which is also called the *infinitesimal generator* of the process. This matrix has the properties of the continuous time (Markov) matrix: $q_{ii} = -q_i = \sum_{i \neq j} q_{ij}$.

$$Q = \begin{Vmatrix} -\lambda_0 & \lambda_0 & 0 & 0 & \dots \\ \mu_1 & -(\lambda_1 + \mu_1) & \lambda_1 & 0 & \dots \\ 0 & \mu_2 & -(\lambda_2 + \mu_2) & \lambda_2 & \dots \\ 0 & 0 & \mu_3 & -(\lambda_3 + \mu_3) & \dots \\ \vdots & \vdots & \vdots & \vdots & \end{Vmatrix}$$

Solving $p_{ij}(h) = \delta_{ij} + hq_{ij} + o(h)$ with $h \searrow 0$ for q_{ij} gives:

$$q_{ij} = \frac{p_{ij}(h) - \delta_{ij}}{h}$$

We can write $\delta_{ij} = p_{ij}(0)$ since we go from state i to state i in zero time steps with probability 1, and from i to j (different from i) in zero time steps with probability 0. Therefore,

$$q_{ij} = \frac{p_{ij}(h) - p_{ij}(0)}{h} = p_{ij}^{'}(0)$$

Here $p_{ij}^{'}(0)$ is a derivative of $p_{ij}(t)$ with respect to t evaluated at 0. Therefore,

$$\mathbf{Q} = [q_{ij}] = [p_{ij}^{'}(0)] = \mathbf{P}^{'}(0)$$

Since $p_{ij}(t)$ are transition probabilities we have:

$$\sum_j p_{ij}(t) = 1$$

Differentiating the above term by term and setting $t = 0$, will give us

$$\sum_j q_{ij} = 0$$

Note that

$$q_{ij} = p_{ij}^{'}(0) \geq 0 \qquad for\ i \neq j$$

$$q_{ii} = p_{ii}^{'}(0) \leq 0$$

which characterizes the infinitesimal transition matrix **Q** described above.

5 Differential Equations of Birth and Death processes

Now, let us move to deriving $P_{ij}(t)$ by using the knowledge about $P_{ij}(h)$. In the case of pure birth and death process (or more generally Continuous time Markov process), the transition probabilities $P_{ij}(t)$ satisfy a system of differential equations known as forward and backward Kolmogorov differential equations.

5.1 Backward Kolmogorov differential equation

The *backward Kolmogorov differential equation* describes the transition probabilities in their dependence on the initial point i. Basically, it analyzes the time interval (0,t+h) by the "first step analysis". It decomposes the (0,t+h) into two intervals (0,h) and (h, t+h), where the first interval h is short and positive.

Formally speaking,

$$
\begin{aligned}
P_{ij}(t+h) &= \sum_{k=0}^{\infty} P_{ik}(h)P_{kj}(t) && (20)\\
&= P_{i,i-1}(h)P_{i-1,j}(t) + P_{i,i+1}(h)P_{i+1,j}(t) + P_{i,i}(h)P_{i,j}(t) && (21)\\
&+ \sum_{k}^{'} P_{ik}(h)P_{kj}(t) && (22)
\end{aligned}
$$

where the last summation is over $k \neq i-1, i, i+1$. By using the facts that $P_{i,i+1}(h) = \lambda_i h + o(h)$, $P_{i,i-1}(h) = \mu_i h + o(h)$, and $P_{i,i}(h) = 1 - h(\lambda_i + \mu_i) + o(h)$, we rewrite

$$
\begin{aligned}
P_{ij}(t+h) &= (\mu_i h + o(h))P_{i-1,j}(t) + (\lambda_i h + o(h))P_{i+1,j}(t) && (23)\\
&+ (1 - h(\lambda_i + \mu_i) + o(h))P_{i,j}(t) + \sum_{k}^{'} P_{ik}(h)P_{kj}(t) && (24)
\end{aligned}
$$

Let us solve for $\sum_{k}^{'} P_{ik}(h)P_{kj}(t)$:

$$
\begin{aligned}
\sum_{k}^{'} P_{ik}(h)P_{kj}(t) &\leq \sum_{k}^{'} P_{ik}(h) && (25)\\
&= 1 - [P_{ii}(h) + P_{i,i-1}(h) + P_{i,i+1}(h)] && (26)\\
&= 1 - [1 - h(\lambda_i + \mu_i) + o(h) + \mu_i h + o(h) + \lambda_i h + o(h)] && (27)\\
&= o(h) && (28)
\end{aligned}
$$

Therefore

$$
\begin{aligned}
P_{ij}(t+h) &= \mu_i h P_{i-1,j}(t) + \lambda_i h P_{i+1,j}(t) + [1 - h(\lambda_i + \mu_i)]P_{i,j}(t) + o(h) && (29)\\
&= \mu_i h P_{i-1,j}(t) + \lambda_i h P_{i+1,j}(t) + P_{i,j}(t) - P_{i,j}(t)h(\lambda_i + \mu_i) + o(h) && (30)
\end{aligned}
$$

Let us transpose $P_{i,j}(t)$ to the right and divide both sides by h, then

$$
\begin{aligned}
\frac{P_{ij}(t+h) - P_{i,j}(t)}{h} &= \frac{\mu_i h P_{i-1,j}(t) + \lambda_i h P_{i+1,j}(t) - P_{i,j}(t)h(\lambda_i + \mu_i) + o(h)}{h} && (31)\\
&= \mu_i P_{i-1,j}(t) + \lambda_i P_{i+1,j}(t) - P_{i,j}(t)(\lambda_i + \mu_i) + o(1) && (32)\\
& && (33)
\end{aligned}
$$

Thus

$$P_{ij}^{'}(t) = \mu_i P_{i-1,j}(t) + \lambda_i P_{i+1,j}(t) - P_{i,j}(t)(\lambda_i + \mu_i)$$

We can now derive a system of differential equations (knowing that $\mu_0 = 0$)

$$
\begin{aligned}
P_{0j}^{'}(t) &= \mu_0 P_{i-1,j}(t) + \lambda_0 P_{0+1,j}(t) - P_{0,j}(t)(\lambda_0 + \mu_0) && (34)\\
&= \lambda_0 P_{1,j}(t) - \lambda_0 P_{0,j}(t) && (35)
\end{aligned}
$$

and

$$P_{ij}^{'}(t) = \mu_i P_{i-1,j}(t) + \lambda_i P_{i+1,j}(t) - P_{i,j}(t)(\lambda_i + \mu_i) \qquad (36)$$

with the initial condition $P_{ij}(0) = \delta_{ij}$.

5.1.1 Using Infinitesimal generator

Similarly, we will now derive a backward Kolmogorov equation by using the matrix notation: since $Q = [q_{ij}] = [p'_{ij}(0)] = P'(0)$ and by Chapman-Kolmogorov equation

$$p_{ij}(s+t) = \sum_k p_{ik}(s)p_{kj}(t)$$

we can differentiate with respect to s so that

$$p'_{ij}(s+t) = \sum_k p'_{ik}(s)p_{kj}(t)$$

Setting $s = 0$ gives

$$p'_{ij}(t) = \sum_k p'_{ik}(0)p_{kj}(t) \tag{37}$$

$$= \sum_k q_{ik}p_{kj}(t) \tag{38}$$

This gives

$$\mathbf{P}'(t) = \mathbf{QP}(t)$$

which defines a ***Kolmogorov backward equation***.

5.2 Forward Kolmogorov differential equation

On the other hand, the ***forward Kolmogorov differential equation*** describes the probability distribution of a state in time t keeping the initial point fixed. It decomposes the time interval (0,t+s) into (0,t) and (t, t+h) by a so-called "last step analysis". Similarly to the backward Kolmogorov differential equation:

$$P_{ij}(t+h) = \sum_{k=0}^{\infty} P_{ik}(t)P_{kj}(h) \tag{39}$$

$$= P_{i,j-1}(t)P_{j-1,j}(h) + P_{i,j+1}(t)P_{j+1,j}(h) + P_{i,j}(t)P_{j,j}(h) \tag{40}$$

$$+ \sum_k{}' P_{ik}(t)P_{kj}(h) \tag{41}$$

The last summation is for $k \neq j-1, j, j+1$. Then,

$$P_{ij}(t+h) = P_{i,j-1}(t)\lambda_{j-1}h + P_{i,j+1}(t)\mu_{j+1}h + P_{i,j}(t)[1 - h(\lambda_j + \mu_j)] + o(h) \tag{42}$$

Similarly to the previous analysis, by translocation $P_{ij}(t)$ and dividing both sides by h, we get two differential equations

$$P'_{ij}(t) = P_{i,j-1}(t)\lambda_{j-1} + P_{i,j+1}(t)\mu_{j+1} - P_{i,j}(t)(\lambda_j + \mu_j)]$$

$$P'_{i0}(t) = P_{i,1}(t)\mu_1 - P_{i,0}(t)\lambda_0$$

with the same initial condition $P_{ij}(0) = \delta_{ij}$.

5.2.1 Using Infinitesimal generator

This time we differentiate with respect t, which gives

$$p^{'}_{ij}(s+t) = \sum_k p_{ik}(s)p^{'}_{kj}(t)$$

Setting $t = 0$ gives

$$p^{'}_{ij}(s) = \sum_k p_{ik}(s)p^{'}_{kj}(0) \tag{43}$$

$$= \sum_k p_{ik}(s)q_{kj} \tag{44}$$

The right-hand side is the sum of the elements in the (i) row of $\mathbf{P}(s)$ multiplied by the the (j) column of Q. Thus

$$\mathbf{P}^{'}(s) = \mathbf{P}(s)\mathbf{Q}$$

which defines a ***Kolmogorov forward equation***. The left-hand side is the (ij) element of $\mathbf{P}^{'}(s)$.

5.3 Exponential method of solving the backward Kolmogorov equation

Let us try to solve the backward Kolmogorov equation ($\mathbf{P}^{'}(t) = \mathbf{QP}(t)$) to obtain the explicit expression for $\mathbf{P}(t)$. Let us use the fact that if

$$p^{'}(t) = qp(t)$$

then

$$p(t) = e^{tq}$$

We can use this fact and the condition that $\mathbf{p}(0) = \mathbf{I}$ to solve for $\mathbf{P}(t)$:

$$\mathbf{P}(t) = e^{\mathbf{Q}t}$$

where the matrix $e^{\mathbf{Q}t}$ is defined by the power series

$$e^{\mathbf{Q}t} = \sum_{n=0}^{\infty} \frac{\mathbf{Q}^n t^n}{n!} = \mathbf{I} + \sum_{n=1}^{\infty} \frac{\mathbf{Q}^n t^n}{n!}$$

6 Application to Poisson process

Let us apply the above equations to general Poisson process with rate λ. The infinitesimal transition rates for Poisson process then are

$$\begin{cases} q_{i,i+1} = \lambda, \\ q_{ii} = -\lambda_i \\ q_{ij} = 0, \qquad j \neq \{i, i+1\} \end{cases}$$

Observe that such a Poisson process models a pure birth process. A transition from i to j in n steps resembles a Bernoulli trial performed n times with the probability of success λ (this is the probability of increase by one, which is essential in transition from i to j).

By induction it follows that

$$q_{ij}^{(n)} = \begin{cases} \binom{n}{j-i}\lambda^{j-i}(-\lambda)^{n-(j-i)}, & if\ 0 \leq j-i \leq n \\ 0, & \text{otherwise} \end{cases}$$

We can re-write

$$\binom{n}{j-i}\lambda^{j-i}(-\lambda)^{n-(j-i)} = \binom{n}{j-i}\lambda^{n}(-1)^{n-(j-i)}$$

Therefore the formula

$$\mathbf{P}(t) = e^{\mathbf{Q}t} = \sum_{n=0}^{\infty}\frac{\mathbf{Q}^n t^n}{n!} = \mathbf{I} + \sum_{n=1}^{\infty}\frac{\mathbf{Q}^n t^n}{n!}$$

gives

$$\begin{aligned}
p_{ij}(t) &= \delta_{ij} + \frac{t^1}{1!}q_{ij} + \frac{t^2}{2!}q_{ij}^2 + \dots &(45)\\
&= \sum_{n=0}^{\infty}\frac{t^n}{n!}q_{ij}^{(n)} &(46)\\
&= \sum_{n=0}^{\infty}\frac{t^n}{n!}\binom{n}{j-i}\lambda^{j-i}(-\lambda)^{n-(j-i)} &(47)\\
&= \sum_{n=0}^{\infty}\frac{t^n}{n!}\frac{n!}{(n-(j-i))!(j-i)!}\lambda^{j-i}(-\lambda)^{n-(j-i)} &(48)\\
&= \sum_{n=0}^{\infty}\frac{t^n}{(n-(j-i))!(j-i)!}\lambda^{j-i}(-\lambda)^{n-(j-i)} &(49)\\
&= \sum_{n=0}^{\infty}\frac{t^{(j-i)}}{(j-i)!}\lambda^{j-i}\frac{t^{(n-(j-i))}}{(n-(j-i))!}(-\lambda)^{n-(j-i)} &(50)\\
&= \frac{(\lambda t)^{(j-i)}}{(j-i)!}\sum_{n=0}^{\infty}\frac{(-\lambda t)^{(n-(j-i))}}{(n-(j-i))!} &(51)\\
& &(52)
\end{aligned}$$

Observe, however, that n cannot start from 0 since we need *at least* $j-i$ transitions to acquire $j-i$ changes. Therefore, n should go from $j-i$ instead. If we define $k = j-i$, then

$$\begin{aligned}
p_{ij}(t) &= \frac{(\lambda t)^{(j-i)}}{(j-i)!}\sum_{n=j-i}^{\infty}\frac{(-\lambda t)^{(n-(j-i))}}{(n-(j-i))!} &(53)\\
&= \frac{(\lambda t)^{(j-i)}}{(j-i)!}\sum_{k=0}^{\infty}\frac{(-\lambda t)^{k}}{k!} &(54)\\
&= \frac{(\lambda t)^{(j-i)}}{(j-i)!}e^{-\lambda t} &(55)
\end{aligned}$$

Thus, the transition probabilities coincide with those defined by the Poisson process. In this case $j-i$ represent the number of events (changes) over the interval of time of length t.

7 Correspondence to Moran process (short)

Moran process can be described as a death-and-birth process. The population size i in the general birth-and-death process corresponds to the number of alleles A_1 in the populations. It follows that λ_i corresponds to the probability of increase in the number of A_1 alleles by one $p_{i,i+1}$, when the A_2 allele is chosen to die and A_1 allele is chosen to reproduce. Similarly, μ_i corresponds to $p_{i,i-1}$, the probability that A_1 is chosen to die while A_2 is chosen to reproduce.

8 Next Lecture

In the next lecture, we will discuss probability of fixation and mean time to absorption in Death-and-Birth process, and their applications to Moran process.

References

[1] Warren Ewens, ”Mathematical Population Genetics”, Second edition, 2004, pp 86-91, 104-109.

[2] Howard Taylor and Samuel Karlin, ”An Introduction to Stochastic Modeling”, Third edition, 1998, pp 267-394.

[3] Sidney Resnick, ”Adventures in Stochastic Processes”, 1992, pp 367-375.

[4] Rabi Bhattacharya, Edward Waymire, ”Stochastic processes with Applications”, pp261-271